基于 HNC 理论的文本情感倾向性分析*

高　歌　罗珺玫　王　宇
(大连理工大学管理与经济学部　大连 116024)

摘要: 【**目的**】构建一种更加科学、准确的评论文本情感倾向性分析方法, 解决网络新词难于计算的问题。【**方法**】利用概念层次网络(HNC)理论的符号对偶性计算情感值, 根据建立的规则为新词确定符号, 利用符号重用降低工作量, 实现对新词的处理。【**结果**】通过对已有成果的分析和改进, 最终得到一套较为完善的情感倾向性分析方法, 并使用真实数据进行实验, 验证了该方法的可行性, 同时也发现了待改进之处。【**局限**】目前仅能对网络短文本进行分析, 且新词的加入需采用人工标注的方式。【**结论**】本文方法可行有效, 为文本情感分析提供了新思路。

关键词: 评论文本　情感分析　概念层次网络(HNC)
分类号: TP391

1　引　言

随着互联网的发展和Web2.0的提出, 网络数据不再仅仅包含结构化、标准化的内容。电子商务网站的兴起和微博、知乎等社区的发展使得用户生成内容(User Generated Content, UGC)数量突增。文本分析正是针对这一问题诞生的, 本文主要关注文本的情感分析。在对电子商务平台的用户评论分析和微博等社区的舆情控制上, 文本情感分析的应用十分广泛。然而文本情感分析也遇到了诸多难点, 通过词形理解概念的方法需要强大的字词库支持, 而字词本身以及字词的含义又会随着人们的使用发生变化。种种问题导致文本情感分析不可能形成一种通用的、规范的方法。本文主要从中文文本入手, 试图摆脱词形的困扰, 深入到概念层, 解析中文文本中包含的褒贬情感, 实现情感值量化的计算。

在中文文本情感分析中, 薛丽敏等[1]采用五元模型进行分析, 即从情感倾向性观点的持有者、倾向性的来源、倾向性的指向、倾向性的立场和倾向性的种类刻画中文文本情感倾向性; 朱嫣岚等[2]使用基于HowNet 的词汇语义倾向计算, 提出基于语义相似度的方法和基于语义相关场的方法; 聂卉等[3]基于HowNet, 提出面向评论效用评估的文本情感特征提取方法; 兰秋军等[4]使用依存句法分析的方法计算金融论坛文本情感倾向性; 何跃等[5]结合话题相关性, 利用基于机器学习改进的情感分类方法对抽取博文的情感极性进行分析; 钟义信[6]提出自然语言理解的全信息方法论; 樊康新[7]提出计算词语情感值的方法。以上方法的不足之处是对于文本情感的理解仍是依赖词形进行处理的, 这与人理解文本的过程不一致, 因此会造成较大的工作量和误差, 而且这种方法对于知识库的完整性要求较高, 对于网络新词、不规范的词语更是无法处理。刘玮楠[8]则利用 HNC 理论, 从 HNC 知识库符号入手实现情感值的计算, 试图通过文本向符号的转换来解决对知识库依赖性过高的问题, 但仍然是简单地从词形转换到符号上, 计算过程没有本质改变。

本文以 HNC 理论[9]为基础, 利用 HNC 理论中概念层次[10]和对偶性[11]原理, 对文本分析不是从形式上入手, 而是注重其概念的表达, 这与人在理解文本内

通讯作者: 王宇, ORCID: 0000-0001-9759-3313, E-mail: ywang@dlut.edu.cn。
*本文系国家自然科学基金重点项目"社会化商务中参与者的信誉与信任机理及交易决策研究"(项目编号: 71431002)的研究成果之一。

容时的过程是一致的。同时从已有的 HNC 知识库和符号的生成规律入手计算文本情感值，提出一种解决新词情感值计算问题的方法。

2 情感值计算

2.1 情感词在 HNC 中的体现

(1) 对偶性

对偶性[11]是指"双重对立并且存在对立统一体"这一特征，情感词的特点是具有对偶的概念，情感极性对应褒贬倾向。通过观察研究发现，在 HNC 概念基元符号中，有着具有这一特征的对偶概念，可以用于情感的褒贬倾向性分析。HNC 的每个概念语义网络在 HNC 知识库中，节点在每个概念语义网络分为高层、中层、底层，其中中层节点反映了概念的基本特性[11]。而研究发现，在涉及情感概念表达的节点中，同一父节点下的中层节点通常是一组对偶的概念集合，且这组概念集合在情感上存在差异，即分别表示情感的褒义和贬义，也说明了它们之间存在对偶关系。

(2) 基础情感值确定

在 HNC 知识库中，首先挑选与情感表达有关的节点和与程度副词有关的节点。经过分析发现，HNC 符号以两种方式定义对偶概念。当节点符号后缀为 ekm/ekn 或 ckm 时(k、m、n 均为数字)，表达一组对偶概念，根据 k 的取值可以判断对偶概念的个数[12]。且每种对偶概念的对偶个数为 2 或 3，对于包含情感的节点来说，若对偶概念个数为 2，则第一个节点为积极意义，第二个节点为消极意义；若对偶概念个数为 3，则第一个节点为积极意义，第二个节点为消极意义，第三个节点通常为处于积极与消极之间的意义，一般同时包含或者同时不包含前面两个概念表达的含义，但其自身表示微弱的贬义情感。这样的关系有利于进行情感基础值的确定。

采用基于 HNC 对偶概念来判断情感极性的方法，一个概念节点从开头到倒数第 4 位均用于表示该概念节点从根节点到叶子节点的路径，对偶性体现在结尾的三个字符上，情感极性则通过最后一个数字表示，例如：

71329 爱恨	7202 面对困难的意志表现
71329e21 爱	7202e75 坚定
71329e22 恨	7202e76 动摇
	7202e77 屈服

根据以上例子可以看出，当 k 值为 2 时，例子中共有两个概念节点，其中 m 值为 1 时表达褒义情感，m 值为 2 时表达贬义情感；当 k 值为 7 时，例子中共有三个概念节点，其中 m 值为 5 时表达褒义情感，m 值为 6 时表达贬义情感，m 值为 7 时也表达偏贬义情感。事实上，在 HNC 知识库中的规律也是这样，即第一个概念为褒义，第二个概念为贬义，第三个概念通常表贬义。

基于 HNC 符号中这种对偶概念的规律，在处理文本时，只需分析相应情感词的概念节点特征，寻找对应的 m 值，就可以判断该情感词的褒贬倾向，得到情感词的初始情感值。

(3) 程度副词权重

在具体计算情感倾向时，将褒义倾向的情感值和贬义倾向的情感值定义为 1 和–1，便得到初始情感值。这种方法将抽象的概念进行量化，实现了文本计算的功能。此外，情感的表达不仅仅由情感词表现，程度副词在表达情感倾向时起着很重要的作用，往往具有加强或者减弱甚至反义的作用，例如"不好看"、"比较好看"和"很好看"在表达情感倾向时存在很大的差异。由于 HNC 知识库中程度副词节点符号的排列关系与其含义联系不大，且表示程度含义的副词较少，本文采用人工定义的方式实现情感值权值的计算。

(4) 最终情感值计算

从计算的角度讲，影响情感值的因素主要有两方面，第一是由情感词决定的情感极性也称为基础情感值，第二是程度副词带来的极性程度增大或缩小。本文的计算过程大致如下：对文本进行预处理，得到主题词和情感词、程度副词。对于某个主题词来说，有 m 个描述它的情感词，这 m 个情感词中的每一个情感词又有 n 个程度副词来描述。对于某一个情感词来说，将其 n 个程度副词带来的程度变化与基础情感值做积，得到这个情感词带来的情感值，再将 m 个情感词的计算结果做和即可得到最终的情感值。

2.2 基础情感值确定

对于采集得到的评论文本，采用基于模式的 Bootstrapping 方法[13]实现主题词、情感词、程度副词的抽取。由于基于模式的方法对词义无任何判断力，因此需要对其进行进一步的确认筛选，同时对于单字和词的处理也有一定的差异性。处理步骤及基础情感值确定步骤如图 1 所示。

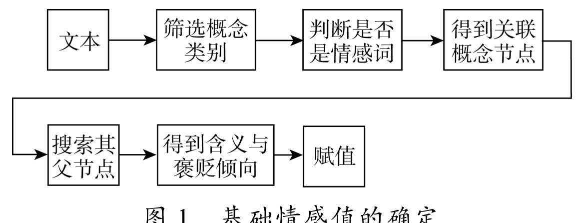


图 1　基础情感值的确定

基于模式的匹配得到的是形式上的情感词，还需进一步确定含义以提高计算精度。对于得到的多字词语，对于每一个词进行如下操作：在 HNC“字词知识库”中查找并筛选概念类别，判断其是否为情感词。判断概念类别，对于某义项概念类别为 u(属性)的词认为可能是情感词。为进一步确认，还需判断其是否具有对偶概念，原则为选取概念类别为 u 的那个义项，查找其关联概念节点或 HNC 符号，包含 ekm、ckm(k、m 分别表述数字)的词即确定为情感词。对于概念类别不是 u 的词语，仍有可能是情感词，处理的方式是找到其关联概念节点后判断是否具有对偶概念，没有对偶概念的词直接舍去，有对偶概念的词则用概念节点反向查找，在词语库中找出同概念节点下概念类别为 u 的词语。如果没有，则舍去该词；如果存在且数目大于设定的阈值则保留该词。在关联概念节点的寻找上，有时会因为某个词语表达含义的层次较低无法在节点库中成功得到，处理的方法是将该关联概念节点从尾部依次减去一个字符进行查找，取能检索到的最长子字符串为其关联概念节点。

通过上述方法对得到的情感词集合再进行处理以得到其基础情感值。取表达了情感倾向的字词关联概念节点，使用关联概念节点在知识库中搜索其父节点，通过上述 HNC 知识库中对偶概念的排列规律赋予初始情感值。对于本身不带有任何程度的情感词，赋予褒义的字词情感值为 1，贬义的为–1，轻度贬义的为–0.5。例如“漂亮”一词，在 HNC 知识库中有两个义项，且两个义项的概念类别都包含 u，进一步解析关联概念节点，三个节点都为对偶概念，且三个对偶概念都为本组对偶概念节点的第一项，因此确定“漂亮”表示的情感基础值为 1；再如“满意”一词，从主观上讲，该词代表了评论者蕴含在其中的褒义情感倾向，但在知识库中“满意”的概念类别为 v，但其 HNC 符号表明该词具有对偶概念，且在同关联概念节点下存在概念类别为 u 的词语，数目通过阈值检验，因此也将其纳入情感词集合，对偶概念识别部分为“e71”，是对偶概念中的第一个，故赋予情感基础值为 1。

对于单字，知识库中没有给出关联概念节点，但可以通过从 HNC 符号中略去概念类别的方式获得，然后对单字进行类似上述方式的处理。例如“好”这个字，很明显包含了评论者主观上褒义的情感。在 HNC 知识库中，“好”的 HNC 符号为“ju51e511”，概念类别为 u，处理后的关联概念节点为“j51e511”，在节点库中找到的关联概念节点为“j51”，而根据“e51”可判断其为对偶概念中的第一个，则赋予情感基础值 1。

2.3　程度副词带来的情感值变动

程度副词对情感值有很大的影响，有必要通过权重定义反映程度副词对其所修饰的情感词产生的影响。

本文的思路是摆脱词形约束，深入到概念层去理解词义。而从概念的角度讲，程度的概念是比较集中、易于分类的。在 HNC 知识库中，概念节点“j60”表示“度的基本内涵”，即与度的表达有关的词语全部与该概念节点相关联。因此无论是什么词，只要包含程度的变化，都会与该节点下的叶子节点关联，根据使用习惯，人工赋予权值：概念节点分为“极、最”、“很”、“稍微”、“适度”、“不够”、“过分”、“否定”，各个节点对应的权值分别为 1.5、1.3、1.1、1、0.8、–0.5、–1。该方法利用 HNC 知识库的概念节点这一概念，减少了数据冗余，简化了计算流程，使分层更加科学，也使计算更加清晰易懂。

在实际计算时，对已经确定基础情感值的文本进行程度副词提取，根据其在 HNC 知识库中关联概念节点的符号确定其给基础值带来的情感值的变动。本文用到的程度副词的概念节点及其权值如表 1 所示。

表 1　程度副词及权值

概念节点	HNC 符号	权值
极、最	j60d01、j60c44	1.5
很	j60c43	1.3
稍微	j60c41	1.1
适度	j60c42、j60e41	1
不够	j60e42	0.8
过分	j60e43	–0.5
否定	!	–1

在 2.2 节得到的基础情感值 $m(w)$上，找到该情感

词附近的程度副词集合，在 HNC 知识库中定位程度副词的概念节点，利用公式(1)计算加权后的情感值，从而将经过加权处理后的情感值作为该情感词的最终情感值。

$$M(w)=(\prod_{o_i \in o} o_i)\times m(w) \qquad (1)$$

其中，$M(w)$表示该情感词加权后的倾向度，o 表示修饰该情感词的程度副词集合，o_i 表示程度副词带来的变动值，$m(w)$表示该情感词的基础情感值。

2.4 网络新词处理

在实际处理中，并不是所有词都能在 HNC 知识库中顺利匹配，有一些网络习惯语、网络新生词并没有及时被包含在知识库中。对于这种情况，提出一种基于 HNC 理论的处理方法。

处理的过程是按照 HNC 理论的规律对新词进行符号确定，对于每个新词，具体的操作如下：人工进行该词的含义辨识，寻找一个与其含义相关的已被知识库收录的词，取该词的 HNC 符号。如果该节点的含义与新词完全一致，则新词与该词符号相同；如果不满足，则寻找含义相近的词的上层概念节点，判断新词是否表达该概念含义。如果可以，则在该节点的下层节点中寻找是否有满足条件的节点，如果有，赋予该新词已有的节点符号；如果没有，为该新词按顺序生成一个新的节点即可，这样就得到了新词的节点符号。另外还需通过词性来确定该词的概念类别，再将概念类别符号与节点符号组合即可得到该词的 HNC 符号，从而确定该词的情感值。

用实例说明上述过程：网络新词“给力”表示说话者对评论客体很满意，或者被评论的客体满足了说话者的需求，且是形容词词形。在 HNC 知识库中，“满意”的关联概念节点是“7123”，“帮助”的关联概念节点是“43e61”。节点“7123”表示的含义是“愿望的实现”，很明显，“给力”表达的即是对事物所期望的结果，因此将“7123”赋给“给力”这个词；同理对于“满意”的关联概念节点“43e61”，表示“支持”的含义，也被收入至“给力”的符号中。从概念类别上看，“给力”属于表达属性的词，即为“u”。

综上，按照该方法生成的 HNC 符号为(u7123/u43e61)，即可按照已建立的方法计算情感值。

新词的确定也是本文尚未解决的问题，留待以后进一步完善。本文实验中对新词进行的处理还是依赖于人工赋予同义词的方法，效率较低。理想的方法是利用已有文本进行自动的同义词识别和情感值的确定。

2.5 最终情感值的计算

在每一条评论中都可能有多个情感词，而每一个情感词也可能被零至多个不同的程度词修饰，所以需要进行统一。在得到基础情感词和程度副词的权值变动后，某条评论的最终情感值计算方法如公式(2)所示。

$$S(w)=\frac{1}{|W|}\sum_{w_i \in W} M(w_i) \qquad (2)$$

其中，$S(w)$表示这条评论中的情感词集合，$|W|$表示情感词的个数，$M(w_i)$表示该情感词加权后的情感值。

3 实验结果

笔者于 2017 年 3 月 3 日通过网页 https://item.jd.com/3650540.html 获取京东自营商品“Apple iPad Air 2 平板电脑 9.7英寸(32G WLAN 版/A8X 芯片/Retina 显示屏/Touch ID 技术 MNV72CH)金色”的评论数据，经过去重、去广告之后，得到共 3 609 条评论。但受制于目前新词处理的低效性，使用前 100 条左右评论产生的时间处于 2016 年 12 月至 2017 年 2 月数据对本方法进行实验，这 100 条左右评论已经足以对该商品目前的反馈进行评估。对数据进行分词、词性标注、主题词抽取、情感词及程度副词抽取以及词频统计等预处理，得到的结果如图 2 所示。

主题词	情感词	程度副词	频次	主题词	情感词	程度副词	频次
商品	划算	-	4	物流	神速	-	1
	耐用	挺	1		及时	-	2
	流畅	-	5			很	1
		很	3		方便	挺	1
	薄	-	1		迅速	-	1
	大	有些	2		慢	比较	1
	好	-	3		厉害	很	1
		太	1		好	-	8
		非常	4			很	5
		相当	1			非常	1
		还	2				
		很	22	包装	精细	-	1
		挺	4		简陋	-	2
	不错	-	30		结实	很	2
		还	11		好	-	2
		很	9			特别	2
		非常	2		不错	-	1
	正常	-	2				
	清晰	-	2	客服	好	-	8
		很	1			很	5
	完美	-	1			非常	1
	合适	-	1		热情	很	2
	便宜	-	4		耐心	-	1

图 2 实验数据

经过对 HNC 知识库的检索，对于知识库中已有的词语使用既定的规律进行赋值。对于知识库中空缺的词语，用上面提到的方法人工生成符号并根据相同的规律赋值。得到的基础情感值如表 2 所示。

表 2　基础情感值

情感词	基础情感值	情感词	基础情感值	情感词	基础情感值
薄	0	精细	1	热情	1
宜	1	厉害	1	神速	1
不错	1	流畅	1	完美	1
大	0	慢	–1	迅速	1
方便	1	耐心	1	正常	1
好	1	耐用	1	及时	1
合适	1	清晰	1	简陋	–1
划算	1	结实	1		

在表 2 基础上，再对程度副词进行处理。使用同样的方法，先在已有的知识库中进行筛选，然后对不存在的词进行人工生成符号并确定程度副词带来的偏差，得到的结果如表 3 所示。

表 3　程度副词带来的变动

程度副词	有些	相当	挺	特别	太	很	还	非常	比较
变动权值	1.1	1.3	1.3	1.3	1.5	1.3	1	1.3	1

对数据进行上述预处理之后，该评论涉及的不同主题情感值计算如下：

$$S(\text{“商品”})=\frac{1}{116}\times\sum_{w_i\in W}M(w_i)=1.083$$

使用本文方法进行实际操作所得的消费者对主题“商品”的情感值为 1.083，略高于 1，所以消费者对商品本身的评价还是趋于积极的。消费者在其他方面的情感值通过相同方法计算得出，结果如下：

$$S(\text{“物流”})=\frac{1}{22}\times\sum_{w_i\in W}M(w_i)=1.123$$

$$S(\text{“包装”})=\frac{1}{8}\times\sum_{w_i\in W}M(w_i)=0.575$$

$$S(\text{“客服”})=\frac{1}{17}\times\sum_{w_i\in W}M(w_i)=1.141$$

同理，计算对商品其他方面评论的情感值。综上，从“商品”、“物流”、“包装”、“客服”这 4 个方面来说，通过得到的数据可知，消费者对商品本身满意度还可以，而物流、客服的情感值则表明这两方面能带给消费者较好的体验，包装的情感值为 0.575，可知在包装上带给消费者的满意度不高。情感值计算的实用意义也就凸显出来，它可以指导该商品的销售方面应该加强包装的质量，从而提升客户的满意度。

4　结　语

在电子商务快速发展、人们网购热情高涨的大环境下，以网购评论为数据基础，应用 HNC 理论的相关概念，对获取的真实评论文本进行计算分析，提出基于 HNC 理论的文本情感倾向性分析方法。在计算情感值时充分利用 HNC 中对偶性概念、概念节点的理论，以及概念层次树挂靠节点所体现出来的新词处理优势，识别用户商品评论中表达出来的情感倾向。

在实验中，首先利用 HNC 中的概念类别对文本中的情感词进行筛选和提取，通过该情感词的关联概念节点找到其父节点，利用对偶性概念赋予基础情感值，然后计算程度副词带来的权值修正，最终对整条评论进行统一计算，得到该商品评论的最终情感值。除此之外，还根据 HNC 概念层次树挂靠节点的理论提出网络新词 HNC 符号生成的规则，对以后基于 HNC 的研究有借鉴意义。

本文主要有以下创新：利用 HNC 概念层次树理论提出通过挂靠节点为网络新词生成 HNC 符号的方法，克服了其他知识库所不能解决的库外词语赋值问题。

虽然提出基于 HNC 的情感倾向性计算方法，实现网购评论量化分析，但是还有很多工作需要进一步的研究：分析的评论和商品较少，下一步将结合更多平台不同类型商品的网购评论进行实验；提出的新词收录的挂靠节点的方法需要人工标注，前期工作量较大，尝试通过机器学习的方法进行改进。

从结果上讲，本实验只是初步验证了方法的可行性。由于情感极性分析本身的高度主观性，实验尚未验证方法的精确性，只能大致通过人工阅读评论的方式确定计算的结果与人为理解的结果一致。同时由于还未实现对新词的自动识别和处理，未进行批量的计算，这也是今后重点要解决的问题。

参考文献：

[1] 薛丽敏，李殿伟，肖斌. 中文文本情感倾向性五元模型研究[J]. 通信技术，2011，44(7)：130-132. (Xue Limin, Li

Dianwei, Xiao Bin. Study on Novel Quintuple Model for Chinese Text Sentiment Orientation [J]. Communications Technology, 2011, 44(7): 130-132.)

[2] 朱嫣岚, 闵锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报, 2006, 20(1): 14-20. (Zhu Yanlan, Min Jin, Zhou Yaqian, et al. Semantic Orientation Computing Based on HowNet [J]. Journal of Chinese Information Processing, 2006, 20(1): 14-20.)

[3] 聂卉, 容哲. 面向评论效用评估的文本情感特征提取[J]. 现代图书情报技术, 2015(7-8): 113-121. (Nie Hui, Rong Zhe. Review Helpfulness Prediction Research Based on Review Sentiment Feature Sets [J]. New Technology of Library and Information Service, 2015(7-8): 113-121.)

[4] 兰秋军, 刘文星, 李卫康, 等. 融合句法信息的金融论坛文本情感计算研究[J]. 现代图书情报技术, 2016(4): 64-71. (Lan Qiujun, Liu Wenxing, Li Weikang, et al. Sentiment Analysis of Financial Forum Textual Message [J]. New Technology of Library and Information Service, 2016(4): 64-71.)

[5] 何跃, 肖敏, 张月. 结合话题相关性的热点话题情感倾向研究[J]. 数据分析与知识发现, 2017, 1(3): 46-53. (He Yue, Xiao Min, Zhang Yue. Sentiment Analysis of Trending Topics Based on Relevance[J]. Data Analysis and Knowledge Discovery, 2017, 1(3): 46-53.)

[6] 钟义信. 自然语言理解的全信息方法论[J]. 北京邮电大学学报, 2004, 27(4): 1-12. (Zhong Yixin. Comprehensive Information Based Methodology for Natural Language Understanding[J]. Journal of Beijing University of Posts and Telecommunications, 2004, 27(4): 1-12.)

[7] 樊康新. 基于多种情感特征的网络文本倾向性判别方法研究[J]. 电脑知识与技术, 2015, 11(22): 18-21. (Fan Kangxin. Research on the Method of Network Text Orientation Discrimination Based on Multiple Sentiment Features[J]. Computer Knowledge and Technology, 2015, 11(22): 18-21.)

[8] 刘玮楠. 基于 HNC 理论的网购评论情感倾向性分析研究[D]. 大连: 大连理工大学, 2013. (Liu Weinan. Research on Sentiment Orientation Analysis of Online-shopping Review Base-on HNC Theory [D]. Dalian: Dalian University of Technology, 2013.)

[9] 黄曾阳. HNC 理论概要[J]. 中文信息学报, 1997, 11(4): 12-21. (Huang Zengyang. The Profile of HNC Theory[J]. Journal of Chinese Information Processing, 1997, 11(4): 12-21.)

[10] 唐兴全. HNC 理论的五元组与词性[C]//自然语言理解与机器翻译——全国第六届计算语言学联合学术会议论文集. 2001. (Tang Xingquan. The Quintuple of HNC Theory and Part of Speech[C]// Proceedings of the 6th Academic Conference on Computational Linguistics in China. 2001.)

[11] 李颖, 池毓焕. 对偶性概念的 HNC 阐释[J]. 中文信息学报, 2004, 18(3): 39-46. (Li Ying, Chi Yuhuan. Re-Categorization of Antithesis Based on HNC Theory[J]. Journal of Chinese Information Processing, 2004, 18(3): 39-46.)

[12] HNC(概念层次网络)理论[C]//中国中文信息学会第六次全国会员代表大会暨成立二十五周年学术会议中文信息处理重大成果汇报展资料汇编. 中国中文信息学会, 2006. (HNC (Hierarchical Network of Concepts) Theory[C]// Proceedings of the 25th Anniversary Academic Conference of Chinese Information Society. 2006.)

[13] 王昌厚, 王菲. 使用基于模式的 Bootstrapping 方法抽取情感词[J]. 计算机工程与应用, 2014, 50(1): 127-129. (Wang Changhou, Wang Fei. Extracting Sentiment Words Using Pattern Based Bootstrapping Method [J]. Computer Engineering and Applications, 2014, 50(1): 127-129.)

作者贡献声明:

高歌: 提出研究思路, 设计研究方案;
罗珺玫: 采集、分析数据;
高歌, 罗珺玫: 进行实验, 起草、修订论文;
王宇: 提出研究问题, 论文最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: 17615002656@163.com。

[1] 高歌, 罗珺玫. 京东评论.xlsx. 情感值计算所用评论.

收稿日期: 2017-05-27
收修改稿日期: 2017-07-03

Analyzing Textual Sentiment Based on HNC Theory

Gao Ge　Luo Junmei　Wang Yu
(Faculty of Management and Economics, Dalian University of Technology, Dalian 116024, China)

Abstract: [**Objective**] This sutdy proposes a new method to conduct sentiment analysis with comment texts, aiming to deal with the issues facing new online terms. [**Methods**] Based on the Hierarchical Network of Concepts (HNC) theory, we defined symbols for the new words, which could be processed more efficiently. [**Results**] The proposed method analyzed the sentiment of the textual message effectively. [**Limitations**] Our method could only process short texts, while we still need to manually create symbols for the new words. [**Conclusions**] We proposed an effective way to conduct sentiment analysis.

Keywords: Comment Text　Sentiment Analysis　Hierarchical Network of Concepts (HNC)

学术研究：领域专家各有所长，但计算机程序能从中做出最优选择

将人类的直觉与计算机的公正性相结合可以改善组织决策，实现降低成本和提高利润。

最近一项研究表明，经过计算机程序的分析，帮助 Dow AgroSciences 公司的业务部门提高了预测的准确性，使其利润增长了 2%至 3%，成本下降了 6%至 7%。美国宾州州立大学 Smeal 商学院供应链管理专业助理教授 Saurabh Bansal 表示。

该团队与 Dow AgroSciences 管理部门的生产专家合作，改进了公司玉米种子部门的预测工作。研究人员指出，受需求变化和天气变化等因素的影响，生产玉米种子可能是一项棘手的工作。“每年，公司都需要弄清楚农民们预计会种植玉米的面积，” Bansal 说：“但是，在这个竞争激烈的行业，经常出现新品种的玉米种子，而且公司在种植新产品方面也缺乏经验，从而无法预知产量。为了优化种植种子的资源配置，就需要对产量进行估算。”

公司经常依靠经理作为专家来对未来事件和活动进行预计，因为这比研究人员实地收集信息更加经济有效。然而，这些根据多年经验所提供的心理模型做出预测的专家往往会引入个人偏见，从而与真实的预测产生偏差。

研究人员开发了计算机模型来估计产量。首先，收集领域专家对产量的判断，然后使用数学模型将分位数估计值转化为产量的平均值和标准偏差。“这个平均值提供了企业平均预期产量的估算值，而标准偏差可以预测增长过程中的预期变化。”

Bansal 表示，将来可以在其他行业实施该模式，包括生物燃料工业和半导体行业，这些行业通常存在严重的供应不确定性。

(编译自: https://www.sciencedaily.com/releases/2017/08/170824110603.htm)

(本刊讯)